

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: SPEAR, Patricia G.
MONTGOMERY, Rebecca I.
- (ii) TITLE OF INVENTION: HERPES VIRUS ENTRY RECEPTOR PROTEIN
- (iii) NUMBER OF SEQUENCES: 7
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: DRESSLER, GOLDSMITH, SHORE & MILNAMOW
 - (B) STREET: 180 N. STETSON, SUITE 4700
 - (C) CITY: CHICAGO
 - (D) STATE: ILLINOIS
 - (E) COUNTRY: U.S.A
 - (F) ZIP: 60601
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (vii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: NORTHRUP, THOMAS E.
 - (B) REGISTRATION NUMBER: 33,268
 - (C) REFERENCE/DOCKET NUMBER: XX
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (312) 616-5400
 - (B) TELEFAX: (312) 616-5460
 - (C) TELEX: XX

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1719 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (ix) FEATURE:
 - (A) NAME/KEY: mat_peptide
 - (B) LOCATION: 293..1189
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 293..1192
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptid
 - (B) LOCATION: 293..406
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCTTCATACC GGCCCTTCCC CTCGGCTTTG CCTGGACAGC TCTGCCTCCC GCAGGGCCCA	60
CCTGTGTCCC CCAGCGCCGC TCCACCCAGC AGGCCTGAGC CCCTCTCTGC TGCCAGACAC	120
CCCCTGCTGC CCACTCTCCT GCTGCTCGGG TTCTGAGGCA CAGCTTGTCA CACCGAGGCG	180
GATTCTCTTT CTCTTTCTCT TCTGGCCAC AGCCGACGCA ATGGCGCTGA GTTCCTCTGC	240
TGGAGTTCAT CCTGCTAGCT GGGTTCCCGA GCTGCCGTC TGAGCCTGAG GC ATG	295
Met 1	
GAG CCT CCT GGA GAC TGG GGG CCT CCT CCC TGG AGA TCC ACC CCC AGA	343
Glu Pro Pro Gly Asp Trp Gly Pro Pro Pro Trp Arg Ser Thr Pro Arg	
5 10 15	
ACC GAC GTC TTG AGG CTG GTG CTG TAT CTC ACC TTC CTG GGA GCC CCC	391
Thr Asp Val Leu Arg Leu Val Leu Tyr Leu Thr Phe Leu Gly Ala Pro	
20 25 30	
TGC TAC GCC CCA GCT CTG CCG TCC TGC AAG GAG GAC GAG TAC CCA GTG	439
Cys Tyr Ala Pro Ala Leu Pro Ser Cys Lys Glu Asp Glu Tyr Pro Val	
35 40 45	
GGC TCC GAG TGC TGC CCC ACG TGC AGT CCA GGT TAT CGT GTG AAG GAG	487
Gly Ser Glu Cys Cys Pro Thr Cys Ser Pro Gly Tyr Arg Val Lys Glu	
50 55 60 65	
GCC TGC GGG GAG CTG ACG GGC ACA GTG TGT GAA CCC TGC CCT CCA GGC	535
Ala Cys Gly Glu Leu Thr Gly Thr Val Cys Glu Pro Cys Pro Pro Gly	
70 75 80	
ACC TAC ATT GCC CAC CTC AAT GGC CTA AGC AAG TGT CTG CAG TGC CAA	583
Thr Tyr Ile Ala His Leu Asn Gly Leu Ser Lys Cys Leu Gln Cys Gln	
85 90 95	
ATG TGT GAC CCA GCC ATG GGC CTG CGC GCG ACG CGG AAC TGC TCC AGG	631
Met Cys Asp Pro Ala Met Gly Leu Arg Ala Thr Arg Asn Cys Ser Arg	
100 105 110	
ACA GAG AAC GCC GTG TGT GGC TGC AGC CCA GGC CAC TTC TGC ATC GTC	679
Thr Glu Asn Ala Val Cys Gly Cys Ser Pro Gly His Phe Cys Ile Val	
115 120 125	
CAG GAC GGG GAC CAC TGC GCC GGT GCC GCC GTT ACG CCA CCT CCA GCC	727
Gln Asp Gly Asp His Cys Ala Gly Ala Ala Val Thr Pro Pro Pro Ala	
130 135 140 145	
CGG GCC AGA GGG TGC AGA AGG GAG GCA CCG AGA GTC AGG ACA CCC TGT	775
Arg Ala Arg Gly Cys Arg Arg Glu Ala Pro Arg Val Arg Thr Pro Cys	
150 155 160	
GTC AGA ACT GCC CCC GGG GAC CTT CTC TCC AAT GGG ACC CTG GAG GAA	823
Val Arg Thr Ala Pro Gly Asp Leu Leu Ser Asn Gly Thr Leu Glu Glu	
165 170 175	
TGT CAG CAC CAG ACC AAG TGC AGC TGG CTG GTG ACG AAG GCC GGA GCT	871
Cys Gln His Gln Thr Lys Cys Ser Trp Leu Val Thr Lys Ala Gly Ala	
180 185 190	
GGG ACC AGC AGC TCC CAC TGG GTA TGG TGG TTT CTC TCA GGG AGC CTC	919
Gly Thr Ser Ser Ser His Trp Val Trp Trp Phe Leu Ser Gly Ser Leu	
195 200 205	
GTC ATC GTC ATT GTT TGC TCC ACA GTT GGC CTA ATC ATA TGT GTG AAA	967
Val Ile Val Ile Val Cys Ser Thr Val Gly Leu Ile Ile Cys Val Lys	
210 215 220 225	
AGA AGA AAG CCA AGG GGT GAT GTA GTC AAG GTG ATC GTC TCC GTC CAG	1015

Arg Arg Lys Pro Arg Gly Asp Val Val Lys Val Ile Val Ser Val Gln	
230 235 240	
CGG AAA AGA CAG GAG GCA GAA GGT GAG GCC ACA GTC ATT GAG GCC CTG	1063
Arg Lys Arg Gln Glu Ala Glu Gly Glu Ala Thr Val Ile Glu Ala Leu	
245 250 255	
CAG GCC CCT CCG GAC GTC ACC ACG GTG GCC GTG AGG AGA CAA TAC CCT	1111
Gln Ala Pro Pro Asp Val Thr Thr Val Ala Val Arg Arg Gln Tyr Pro	
260 265 270	
CAT TCA CGG GGA GGA GCC CAA ACC ACT GAC CCA CAG ACT CTG CAC CCC	1159
His Ser Arg Gly Gly Ala Gln Thr Thr Asp Pro Gln Thr Leu His Pro	
275 280 285	
GAC GCC AGA GAT ACC TGG AGC GAC GGC TGC TGA AAGAGGCTGT CCACCTGGCG	1212
Asp Ala Arg Asp Thr Trp Ser Asp Gly Cys *	
290 295 300	
AAACCACCGG AGCCCGGAGG CTTGGGGGCT CCGCCCTGGG CTGGCTTCCG TCTCCTCCAG	1272
TGGAGGGAGA GGTGGGGCCC CTGCTGGGGT AGAGCTGGGG ACGCCACGTG CCATTCCCAT	1332
GGGCCAGTGA GGGCCTGGGG CCTCTGTTCT GCTGTGGCCT GAGCTCCCCA GAGTCCTGAG	1392
GAGGAGCGCC AGTTGCCCCT CGCTCACAGA CCACACACCC AGCCCTCCTG GGCCAGCCCA	1452
GAGGGCCCTT CAGACCCCAG CTGTCTGCGC GTCTGACTCT TGTGGCCTCA GCAGGACAGG	1512
CCCCGGGCAC TGCCTCACAG CCAAGGCTGG ACTGGGTTGG CTGCAGTGTG GTGTTTAGTG	1572
GATACCACAT CGGAAGTGAT TTTCTAAATT GGATTGGAAT TCCGGTCCTG TCTTCTATTT	1632
GTCATGAAAC AGTGTATTTG GGGAGATGCT GTGGGAGGAT GTAAATATCT TGTTTCTCCT	1692
CAAAAAAAAA AAAAAAAAAA AAAAAAA	1719

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 299 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Glu	Pro	Pro	Gly	Asp	Trp	Gly	Pro	Pro	Pro	Trp	Arg	Ser	Thr	Pro
1				5				10						15	
Arg	Thr	Asp	Val	Leu	Arg	Leu	Val	Leu	Tyr	Leu	Thr	Phe	Leu	Gly	Ala
			20					25					30		
Pro	Cys	Tyr	Ala	Pro	Ala	Leu	Pro	Ser	Cys	Lys	Glu	Asp	Glu	Tyr	Pro
			35				40					45			
Val	Gly	Ser	Glu	Cys	Cys	Pro	Thr	Cys	Ser	Pro	Gly	Tyr	Arg	Val	Lys
	50					55					60				
Glu	Ala	Cys	Gly	Glu	Leu	Thr	Gly	Thr	Val	Cys	Glu	Pro	Cys	Pro	Pro
	65				70					75					80
Gly	Thr	Tyr	Ile	Ala	His	Leu	Asn	Gly	Leu	Ser	Lys	Cys	Leu	Gln	Cys
			85				90							95	
Gln	Met	Cys	Asp	Pro	Ala	Met	Gly	Leu	Arg	Ala	Thr	Arg	Asn	Cy	Ser
			100				105						110		

Arg Thr Glu Asn Ala Val Cys Gly Cys Ser Pro Gly His Phe Cys Il
 115 120 125
 Val Gln A p Gly Asp His Cys Ala Gly Ala Ala Val Thr Pro Pro Pro
 130 135 140
 Ala Arg Ala Arg Gly Cys Arg Arg Glu Ala Pro Arg Val Arg Thr Pro
 145 150 155 160
 Cys Val Arg Thr Ala Pro Gly Asp Leu Leu Ser Asn Gly Thr Leu Glu
 165 170 175
 Glu Cys Gln His Gln Thr Lys Cys Ser Trp Leu Val Thr Lys Ala Gly
 180 185 190
 Ala Gly Thr Ser Ser Ser His Trp Val Trp Trp Phe Leu Ser Gly Ser
 195 200 205
 Leu Val Ile Val Ile Val Cys Ser Thr Val Gly Leu Ile Ile Cys Val
 210 215 220
 Lys Arg Arg Lys Pro Arg Gly Asp Val Val Lys Val Ile Val Ser Val
 225 230 235 240
 Gln Arg Lys Arg Gln Glu Ala Glu Gly Glu Ala Thr Val Ile Glu Ala
 245 250 255
 Leu Gln Ala Pro Pro Asp Val Thr Thr Val Ala Val Arg Arg Gln Tyr
 260 265 270
 Pro His Ser Arg Gly Gly Ala Gln Thr Thr Asp Pro Gln Thr Leu His
 275 280 285
 Pro Asp Ala Arg Asp Thr Trp Ser Asp Gly Cys
 290 295

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AACCCGGCTC GAGCGGCCGC T

21

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAATTCACC AACTTAAGG TG

22

(2) INFORMATION FOR SEQ ID NO:5:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ACAAGACCGT TGCACCTC

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4619 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 64..1320

(ix) FEATURE:
 (A) NAME/KEY: mat_peptide
 (B) LOCATION: 64..1317

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AAGCTTGCA	TGCCTGCAGGT	CGACTCTAGC	TGGGTTCCCG	AGCTGCCGGT	CTGAGCCTGA	60
GGC ATG GAG CCT CCT GGA GAC TGG GGG CCT CCT CCC TGG AGA TCC ACC	108					
Met Glu Pro Pro Gly Asp Trp Gly Pro Pro Pro Trp Arg Ser Thr						
1 5 10 15						
CCC AGA ACC GAC GTC TTG AGG CTG GTG CTG TAT CTC ACC TTC CTG GGA	156					
Pro Arg Thr Asp Val Leu Arg Leu Val Leu Tyr Leu Thr Phe Leu Gly						
20 25 30						
GCC CCC TGC TAC GCC CCA GCT CTG CCG TCC TGC AAG GAG GAC GAG TAC	204					
Ala Pro Cys Tyr Ala Pro Ala Leu Pro Ser Cys Lys Glu Asp Glu Tyr						
35 40 45						
CCA GTG GGC TCC GAG TGC TGC CCC ACG TGC AGT CCA GGT TAT CGT GTG	252					
Pro Val Gly Ser Glu Cys Cys Pro Thr Cys Ser Pro Gly Tyr Arg Val						
50 55 60						
AAG GAG GCC TGC GGG GAG CTG ACG GGC ACA GTG TGT GAA CCC TGC CCT	300					
Lys Glu Ala Cys Gly Glu Leu Thr Gly Thr Val Cys Glu Pro Cys Pro						
65 70 75						
CCA GGC ACC TAC ATT GCC CAC CTC AAT GGC CTA AGC AAG TGT CTG CAG	348					
Pro Gly Thr Tyr Ile Ala His Leu Asn Gly Leu Ser Lys Cys Leu Gln						
80 85 90 95						
TGC CAA ATG TGT GAC CCA GCC ATG GGC CTG CGC GCG ACG CGG AAC TGC	396					
Cys Gln Met Cys Asp Pro Ala Met Gly Leu Arg Ala Thr Arg Asn Cys						
100 105 110						
TCC AGG ACA GAG AAC GCC GTG TGT GGC TGC AGC CCA GGC CAC TTC TGC	444					
Ser Arg Thr Glu Asn Ala Val Cys Gly Cys Ser Pro Gly His Phe Cys						
115 120 125						
ATC GTC CAG GAC GGG GAC CAC TGC GCC GGT GCC GCC GTT ACG CCA CCT	492					
Ile Val Gln Asp Gly Asp His Cys Ala Gly Ala Ala Val Thr Pro Pro						

130	135	140	
CCA GCC CGG GCC AGA GGG TGC AGA AGG GAG GCA CCG AGA GTC AGG ACA Pro Ala Arg Ala Arg Gly Cys Arg Arg Glu Ala Pr Arg Val Arg Thr 145 150 155	540		
CCC TGT GTC AGA ACT GCC CCC GGG GAC CTT CTC TCC AAT GGG ACC CTG Pro Cys Val Arg Thr Ala Pro Gly Asp Leu Leu Ser Asn Gly Thr Leu 160 165 170 175	588		
GAG GAA TGT CAG CAC CAG ACC AAG TGC AGA ATT CAC AAG ACC GTT GCA Glu Glu Cys Gln His Gln Thr Lys Cys Arg Ile His Lys Thr Val Ala 180 185 190	636		
CCC TCG ACA TGC AGC AAG CCC ACG TGC CCA CCC CCT GAA CTC CTG GGG Pro Ser Thr Cys Ser Lys Pro Thr Cys Pro Pro Pro Glu Leu Leu Gly 195 200 205	684		
GGA CCG TCT GTC TTC ATC TTC CCC CCA AAA CCC AAG GAC ACC CTC ATG Gly Pro Ser Val Phe Ile Phe Pro Pro Lys Pro Lys Asp Thr Leu Met 210 215 220	732		
ATC TCA CGC ACC CCC GAG GTC ACA TGC GTG GTG GTG GAC GTG AGC CAG Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser Gln 225 230 235	780		
GAT GAC CCC GAG GTG CAG TTC ACA TGG TAC ATA AAC AAC GAG CAG GTG Asp Asp Pro Glu Val Gln Phe Thr Trp Tyr Ile Asn Asn Glu Gln Val 240 245 250 255	828		
CGC ACC GCC CGG CCG CCG CTA CGG GAG CAG CAG TTC AAC AGC ACG ATC Arg Thr Ala Arg Pro Pro Leu Arg Glu Gln Gln Phe Asn Ser Thr Ile 260 265 270	876		
CGC GTG GTC AGC ACC CTC CCC ATC ACG CAC CAG GAC TGG CTG AGG GGC Arg Val Val Ser Thr Leu Pro Ile Thr His Gln Asp Trp Leu Arg Gly 275 280 285	924		
AAG GAG TTC AAG TGC AAA GTC CAC AAC AAG GCA CTC CCG GCC CCC ATC Lys Glu Phe Lys Cys Lys Val His Asn Lys Ala Leu Pro Ala Pro Ile 290 295 300	972		
GAG AAA ACC ATC TCC AAA GCC AGA GGG CAG CCC CTG GAG CCG AAG GTC Glu Lys Thr Ile Ser Lys Ala Arg Gly Gln Pro Leu Glu Pro Lys Val 305 310 315	1020		
TAC ACC ATG GGC CCT CCC CGG GAG GAG CTG AGC AGC AGG TCG GTC AGC Tyr Thr Met Gly Pro Pro Arg Glu Glu Leu Ser Ser Arg Ser Val Ser 320 325 330 335	1068		
CTG ACC TGC ATG ATC AAC GGC TTC TAC CCT TCC GAC ATC TCG GTG GAG Leu Thr Cys Met Ile Asn Gly Phe Tyr Pro Ser Asp Ile Ser Val Glu 340 345 350	1116		
TGG GAG AAG AAC GGG AAG GCA GAG GAC AAC TAC AAG ACC ACG CCG GCC Trp Glu Lys Asn Gly Lys Ala Glu Asp Asn Tyr Lys Thr Thr Pro Ala 355 360 365	1164		
GTG CTG GAC AGC GAC GGC TCC TAC TTC CTC TAC AAC AAG CTC TCA GTG Val Leu Asp Ser Asp Gly Ser Tyr Phe Leu Tyr Asn Lys Leu Ser Val 370 375 380	1212		
CCC ACG AGT GAG TGG CAG CGG GGC GAC GTC TTC ACC TGC TCC GTG ATG Pro Thr Ser Glu Trp Gln Arg Gly Asp Val Phe Thr Cys Ser Val Met 385 390 395	1260		
CAC GAG GCC TTG CAC AAC CAC TAC ACG CAG AAG TCC ATC TCC CGC TCT His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Ile Ser Arg Ser 400 405 410 415	1308		

CCG GGT AAA TGA GCGCTGTGCC GGCGAGCTGC CCCTCTCCCT CCCCCCAGC	1360
Pro Gly Lys *	
CCGCAGCTGT GCACCCCGCA CACAAATAAA GCACCCAGCT CTGCCCTGAA CAGCTTCCGG	1420
TCTCCCTATA GTGAGTCGTA TTAATTTCTGA TAAGCCAGCT GCATTAATGA ATCGGCCAAC	1480
GCGCGGGGAG AGGCGGTTTG CGTATTGGGC GCTCTTCCGC TTCCTCGCTC ACTGACTCGC	1540
TGCGCTCGGT CGTTCGGCTG CGGCGAGCGG TATCAGCTCA CTCAAAGGCG GTAATACGGT	1600
TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGGC CAGCAAAAGG	1660
CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCTGACG	1720
AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT	1780
ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC TGTTCCGACC CTGCCGCTTA	1840
CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC GCTTTCTCAT AGCTCACGGT	1900
GTAGGTATCT CAGTTCGGTG TAGGTCGTTT GCTCCAAGCT GGGCTGTGTG CACGAACCCC	1960
CCGTTTCAGCC CGACCGCTGC GCCTTATCCG GTAACATCG TCTTGAGTCC AACCCGGTAA	2020
GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG	2080
TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT AGAAGGACAG	2140
TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT GGTAGCTCTT	2200
GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTTT TGTGCAAG CAGCAGATTA	2260
CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC	2320
AGTGGAACGA AAATCAGCT TAAGGGATTT TGGTCATGAG ATTATCAAAA AGGATCTTCA	2380
CCTAGATCCT TTAAATTA AAATGAAGTT TAAATCAAT CTAAAGTATA TATGAGTAAA	2440
CTTGGTCTGA CAGTTACCA TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT	2500
TTCGTTCATC CATAGTTGCC TGAATCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT	2560
TACCATCTGG CCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCAGATT	2620
TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT	2680
CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGAAGCTAG AGTAAGTAGT TCGCCAGTTA	2740
ATAGTTTGCG CAACGTTGTT GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTTG	2800
GTATGGCTTC ATTCAGCTCC GGTCCCAAC GATCAAGGCG AGTTACATGA TCCCCATGT	2860
TGTGCAAAA AGCGGTTAGC TCCTTCGGTC CTCGATCGT TGTCAGAAGT AAGTTGGCCG	2920
CAGTGTATC ACTCATGGTT ATGGCAGCAC TGCATAATC TCTTACTGTC ATGCCATCCG	2980
TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC	3040
GGCGACCGAG TTGCTCTTGC CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA	3100
CTTTAAAAGT GTCATCATT GGAAAACGTT CTTGGGGCG AAAACTCTCA AGGATCTTAC	3160
CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT	3220
TTACTTTTAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC GCAAAAAGG	3280
GAATAAGGGC GACACGGAAA TGTGAATAC TCATACTCTT CCTTTTCAA TATTATTGAA	3340

GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA	3400
AACAAATAGG GGTTCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA	3460
TTATTATCAT GACATTAACC TATAAAATA GCGGTATCAC GAGGCCCTTT CGTCTCGCGC	3520
GTTTCGGTGA TGACGGTGAA AACCTCTGAC ACATGCAGCT CCCGGAGACG GTCACAGCTT	3580
GTCTGTAAGC GGATGCCGGG AGCAGACAAG CCCGTCAGGG CGCGTCAGCG GGTGTTGGCG	3640
GGTGTGCGGG CTGGCTTAAC TATGCGGCAT CAGAGCAGAT TGTACTGAGA GTGCACCATA	3700
TCGACGCTCT CCCTTATGCG ACTCCTGCAT TAGGAAGCAG CCCAGTAGTA GGTGAGGCC	3760
GTTGAGCACC GCCGCCGCAA GGAATGGTGC AAGGAGATGG CGCCCAACAG TCCCCGGCC	3820
ACGGGGCCTG CCACCATACC CACGCCGAAA CAAGCGCTCA TGAGCCCGAA GTGGCGAGCC	3880
CGATCTTCCC CATCGGTGAT GTCGGCGATA TAGGCGCCAG CAACCGCACC TGTGGCGCCG	3940
GTGATGCCGG CCACGATGCG TCCGGCGTAG AGGATCTGGC TAGTTATTAA TAGTAATCAA	4000
TTACGGGGTC ATTAGTTCAT AGCCCATATA TGGAGTTCCG CGTTACATAA CTTACGGTAA	4060
ATGGCCCGCC TGGCTGACCG CCCAACGACC CCCGCCATT GACGTCAATA ATGACGTATG	4120
TTCCCATAGT AACGCCAATA GGGACTTTCC ATTGACGTCA ATGGGTGGAC TATTACGGT	4180
AAACTGCCCA CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCC CCTATTGACG	4240
TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCAGTA CATGACCTTA TGGGACTTTC	4300
CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC CATGGTGATG CGGTTTTGGC	4360
AGTACATCAA TGGGCGTGGA TAGCGGTTTG ACTCACGGG ATTTCCAAGT CTCCACCCCA	4420
TTGACGTCAA TGGGAGTTTG TTTTGGCACC AAAATCAACG GGACTTTCCA AAATGTCGTA	4480
ACAACCTCCG CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG GTCTATATAA	4540
GCAGAGCTCT CTGGCTAACT AGAGAACCCA CTGCTTAACT GGCTTATCGA AATTAATACG	4600
ACTCACTATA GGGAGACCC	4619

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 418 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met	Glu	Pro	Pro	Gly	Asp	Trp	Gly	Pro	Pro	Pro	Trp	Arg	Ser	Thr	Pro	1	5	10	15
Arg	Thr	Asp	Val	Leu	Arg	Leu	Val	Leu	Tyr	Leu	Thr	Phe	Leu	Gly	Ala	20	25	30	
Pro	Cys	Tyr	Ala	Pro	Ala	Leu	Pro	Ser	Cys	Lys	Glu	Asp	Glu	Tyr	Pro	35	40	45	
Val	Gly	Ser	Glu	Cys	Cys	Pro	Thr	Cys	Ser	Pro	Gly	Tyr	Arg	Val	Lys	50	55	60	
Glu	Ala	Cys	Gly	Glu	Leu	Thr	Gly	Thr	Val	Cys	Glu	Pro	Cys	Pro	Pro				

65	70	75	80
Gly Thr Tyr Ile Ala His Leu Asn Gly Leu S r Lys Cys Leu Gln Cys	85	90	95
Gln Met Cys Asp Pro Ala Met Gly Leu Arg Ala Thr Arg Asn Cys Ser	100	105	110
Arg Thr Glu Asn Ala Val Cys Gly Cys Ser Pro Gly His Phe Cys Ile	115	120	125
Val Gln Asp Gly Asp His Cys Ala Gly Ala Ala Val Thr Pro Pro Pro	130	135	140
Ala Arg Ala Arg Gly Cys Arg Arg Glu Ala Pro Arg Val Arg Thr Pro	145	150	155
Cys Val Arg Thr Ala Pro Gly Asp Leu Leu Ser Asn Gly Thr Leu Glu	165	170	175
Glu Cys Gln His Gln Thr Lys Cys Arg Ile His Lys Thr Val Ala Pro	180	185	190
S r Thr Cys Ser Lys Pro Thr Cys Pro Pro Pro Glu Leu Leu Gly Gly	195	200	205
Pr Ser Val Phe Ile Phe Pro Pro Lys Pro Lys Asp Thr Leu Met Ile	210	215	220
Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser Gln Asp	225	230	235
Asp Pro Glu Val Gln Phe Thr Trp Tyr Ile Asn Asn Glu Gln Val Arg	245	250	255
Thr Ala Arg Pro Pro Leu Arg Glu Gln Gln Phe Asn Ser Thr Ile Arg	260	265	270
Val Val Ser Thr Leu Pro Ile Thr His Gln Asp Trp Leu Arg Gly Lys	275	280	285
Glu Phe Lys Cys Lys Val His Asn Lys Ala Leu Pro Ala Pro Ile Glu	290	295	300
Lys Thr Ile Ser Lys Ala Arg Gly Gln Pro Leu Glu Pro Lys Val Tyr	305	310	315
Thr Met Gly Pro Pro Arg Glu Glu Leu Ser Ser Arg Ser Val Ser Leu	325	330	335
Thr Cys Met Ile Asn Gly Phe Tyr Pro Ser Asp Ile Ser Val Glu Trp	340	345	350
Glu Lys Asn Gly Lys Ala Glu Asp Asn Tyr Lys Thr Thr Pro Ala Val	355	360	365
Leu Asp Ser Asp Gly Ser Tyr Phe Leu Tyr Asn Lys Leu Ser Val Pro	370	375	380
Thr Ser Glu Trp Gln Arg Gly Asp Val Phe Thr Cys Ser Val Met His	385	390	395
Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Ile Ser Arg Ser Pro	405	410	415
Gly Lys			